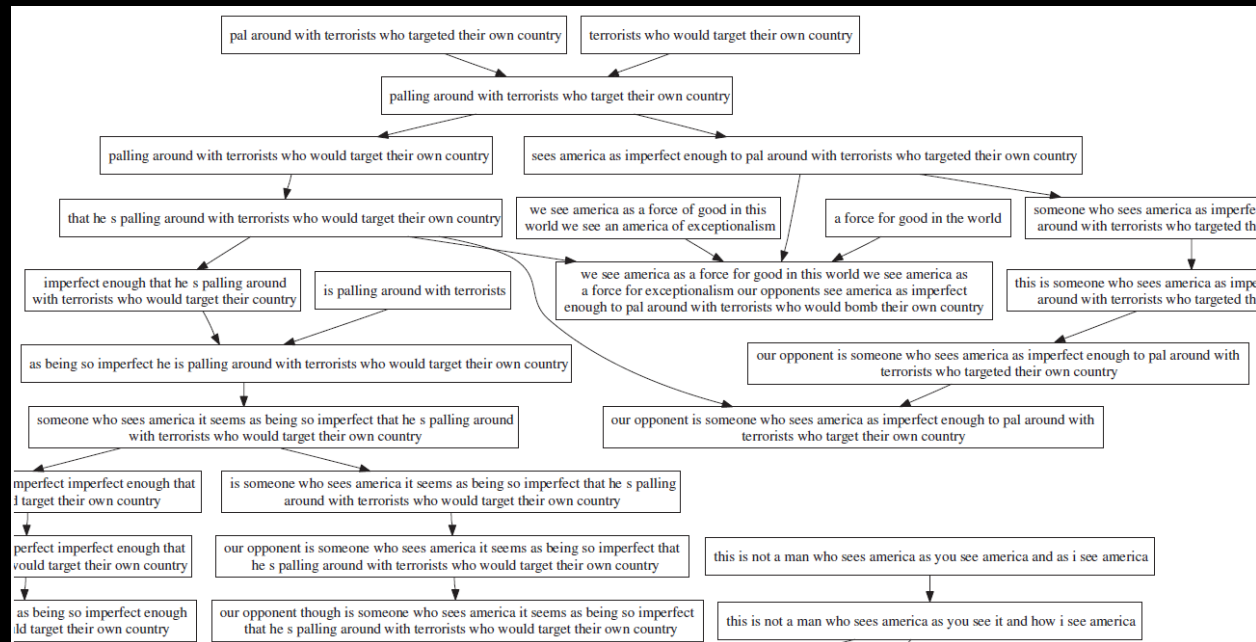


# REACTION Update

## Task 4: Tracking Information Flow



Matt Lease

School of Information

University of Texas at Austin

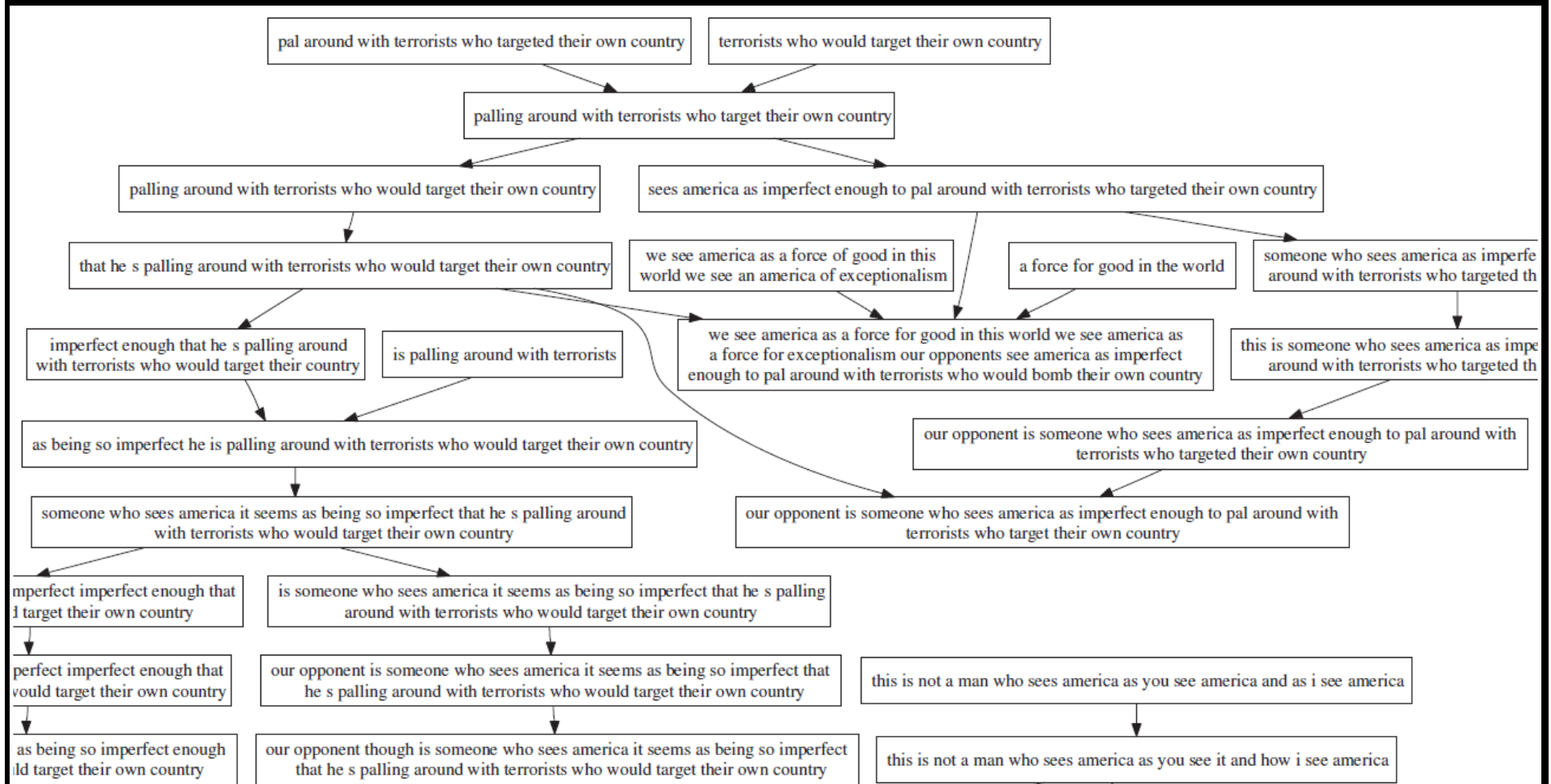
ml@ischool.utexas.edu



# Goal

- Find *memes* in online news/blogosphere
  - Short, similar text snippets shared across sources
  - Inspired by MemeTracker:  
<http://memetracker.org>
  - E.g. “lipstick on a pig” by Palin in 2008 US election
- Once found, could analyze path traveled, edit history, individuals/communities involved, etc.
  - Not there....

# Example (Leskovec et al, KDD'09)



# Prior Work & Our Approach

- Leskovec et al., KDD'09: quotations only
- Steven Skiena, Stony Brook NY
  - Named-entities only: <http://www.textmap.com>
- Kolak and Schilit'08: scanned books
  - Detect memes in arbitrary text (no “” required)
  - MapReduce “shingling” approach
    - N-gram collection, then “grouping” to merge
  - *We adopt this approach...*

# Data

- TREC Blogs08 Collection
  - [http://ir.dcs.gla.ac.uk/test\\_collections/blogs08info.html](http://ir.dcs.gla.ac.uk/test_collections/blogs08info.html)
  - 28M permalinks (January 2008 – January 2009)
  - 250G compressed
- ICWSM 2009 Spinn3r Blog Dataset
  - <http://www.icwsm.org/data/>
  - 44 million blog posts (August - September, 2008)
  - 27 GB compressed

# MapReduce @ UT

- No prior UT MapReduce/Hadoop infrastructure
- UT LIFT Award to Lease et al. in Sept. 2010
- New harddisks @ [TACC](#) installed Dec. 2010
  - 48 Dell R610 nodes
    - 2 Intel Nehalem quad-core processors (8 cores) @ 2.53 GHz
    - 48GB RAM, now with ~1.5TB per node
  - 16 Dell R710 (same CPU configuration)
    - 144GB RAM, now with ~0.8TB available per node
- Setup Hadoop, testing, benchmarking, etc.
- Problems with contention from other users...

# REACTION Funded Personnel

- Hohyon Ryu
  - Entered UT iSchool PhD program in Sept. 2010
    - <http://www.ischool.utexas.edu/~hr>
    - Completed two semesters (9 funded months)
      - took 6 courses + REACTION work + setup Hadoop on TACC
    - On internship at [InfoChimps](#) since May, returns in Aug.
      - REACTION to fund Sept. 2011 – May 2012 (likely internship again in 2012)



# Meme Finding Demo:

<http://stingray.ischool.utexas.edu/index.php?>

Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://stingray.ischool.utexas.edu/index.php?meme\_id=127564

Most Visited me Dates Wiki accounts transport class austin papers IR Conf Tech pics Grant fab Meetup NPR cook Subscribe Most Visited

Google ... Academ... Editoria... Panos I... Parbash... Amazon... Amazon... Submit ... Quick St... Cloud9: ... Cloud9: ... Program Google ... Data Br... ht...64 x firefox ...

28

- children who we love with 28
- five wonderful children who we 28
- former mayor of a town 28
- heart and mean everything to 28
- i don t know if 28
- love with all our heart 28
- of like a community

five wonderful children who we

Related Docs

<http://advanceindiana.blogspot.com/2008/09/palin-discloses-17-year-old-daughters.html>

<http://wolflowling.blogspot.com/2008/09/bristols-got-bun.html>

<http://victoriakos.blogspot.com/2008/09/palin-says-daughter-17-pregnant.html>

<http://forestnymph.blogspot.com/2008/09/palins-daughter-17-choose-life-bristol.html>

<http://senescence.blogspot.com/2008/09/governor-palin-daughter.html>

<http://www.reliablepolitics.com>

- with news that as parents
- wonderful children who we love
- would make her grow up

we have been blessed with five wonderful children who we love with all our heart and mean everything to us our beautiful daughter bristol came to us with news that as parents we knew would make her grow up faster than we had ever planned we re proud of bristol s decision to have her baby and even prouder to become grandparents as bristol faces the responsibilities of adulthood she knows she has our unconditional love and support


MARX & THE COMMUNIST MANIFESTO, 21

Feb. 1848

The Palin camp announced this morning that Gov. Palin's eldest daughter, 17 year old Bristol, is pregnant.

SATURDAY ART & VERSE POSTS

gaspirz.com




21 May - Stars: Vermeer, Van Dyke,

From Hot Air, this is the statement released by Gov. Palin and her husband:

"We have been blessed with five wonderful children who we love with all our heart and mean everything to us. Our beautiful daughter Bristol came to us with news that as parents we knew would make her grow up faster than we had ever planned. We're proud of Bristol's decision to have her baby and even prouder to become grandparents. As Bristol faces the responsibilities of adulthood, she knows she has our unconditional love and support.

"Bristol and the young man she will marry are going to realize very quickly the difficulties of raising a child, which is why they will have the love and support of our entire family. We ask the media to respect our daughter and Levi's privacy as has always been the tradition of children of candidates."

WATCHER'S COUNCIL AL



American Digest

Cheat Seaking Missles

Hillbilly White Trash

The Education Wonks

The Provocateur



# Steve Berry's Course Project: Crowdsourcing News Annotation

Annotating article 73

[Return to article list](#)

## Finally, the train intrigues this Texan

Source: bizj

Original Article: <http://www.bizjournals.com/austin/stories/2010/01/04/editorial1.html>

Hell must have frozen over and any minute now I expect a pig to fly past my office window.

Why? Because I — a stubborn Texan in love with my car and who can count on one hand the number of times I've used the public transit system here in the past 20 years — have a tinge of excitement now for our new commuter train. Yes, despite its ridiculous delays and no start date yet.

There are two reasons for this newfound — and arguably overdue — enthusiasm.

First, I recently got a reminder from BART how well trains can move people within a metropolitan area. Sure, I've ridden commuter rails — from New York to Los Angeles — many times, but on a recent trip to San Francisco I spent more time on the Bay Area Rapid Transit trains than ever before, and evaluated the system for the first time with a scrutinous eye because our rail line is, of course, the talk of the town.

And while the dynamics of the Bay Area drastically differ from Central Texas' — as do the rail lines themselves — it ignited a hope and expectation that the rail here can work. Because it certainly works there. And being the selfish person I am, now that I've benefited from it personally, it qualifies to me as a worthwhile investment for that community.

But the other reason why I'm now intrigued about riding Austin's new rails revolves around a move I've been pondering from South Austin to North Austin. I've lived in Far South Austin for about a decade now, so all this talk about rail has never thrilled me. There's never been serious talk of a rail line heading south of downtown, so again — looking at things through a self-centered lens — my tax dollars were going to something that I never expected to use.

'Train, schmain' was my philosophy.

But if I move north — especially beyond U.S. Highway 183 — you'd better bet I'll hop on the Red Line at least now and then. Perhaps regularly, I've been thinking lately. I'm even starting to consider homes I previously deemed too far from downtown due to the new rail stations.

This is drastic thinking for a stereotypical Texan who drives his car to visit his neighbor two streets over. And yes, I've even driven my car the 100 yards or so between my front door and mailbox once or twice.

So all you public-transit advocates, take heart. If this change in mindset can happen to me, it can happen to anyone.

I just hope the train starts chugging before cars start flying, because when that happens you'll all have to go back to trying to pry my car keys from my hands.



Total  
Points

0011

@numeralsix

Current Rank: **Adept**

Next rank in 29 points [notifier](#)

## Edit Entities

[Add Entity](#)

Topic



Other



Person



Ive



Social Tag



Rail transport in the United States



Commuter rail



Transportation in the United States



Rail transport



Train



Metro-North Railroad

# Progress

- Meme Detection
  - Have initial output, simple interface demo
  - Need: better false positive removal
    - “splogs”
    - advertisements vs. main text
  - Paper submissions
    - SIGIR’11 poster submitted but rejected
- Simple crowdsourcing annotation demo
- Supporting infrastructure: Hadoop @ UT TACC

# Outlook

- REACTION funding to UT ends Aug 31, 2012
  - Fund Ryu Sept. 2011 – May 2012
    - Likely internship again summer 2012
  - Should be partial funding for another student
- More student projects?
  - Teaching MapReduce & IR courses in Fall'11
  - Teaching Crowdsourcing Spring'12
- Continue focus on meme detection processing
  - possible consumption and/or annotation interfaces for public, journalists, and turkers